

Application No. 09/747,515
Amendment Dated June 1, 2004
Reply to Office Action of April 1, 2004

Remarks

This amendment is responsive to the Office action mailed April 1, 2004 in connection with the above-identified patent application. In that Action, claims 55-59, 61-65, and 67-71 were rejected under 35 U.S.C. § 103(a) as being unpatentable over U.S. Patent No. 5,727,199 to Chen, et al. in view of U.S. Patent No. 5,647,058 to Agrawal, et al.

Claims 60, 66, and 72 were rejected under 35 U.S.C. § 103(a) as being unpatentable over Chen, et al. in view of Agrawal, et al. [1] and further in view of U.S. Patent No. 6,094,651 to Agrawal, et al. [2].

Claims 73-76, 78-81, and 83-86 were rejected under 35 U.S.C. § 103(a) as being unpatentable over Chen, et al. in view of Agrawal, et al. [1] and further in view of Information Builders Inc.

Lastly, claims 77, 82, and 87 were rejected under 35 U.S.C. § 103(a) as being unpatentable over Chen, et al. in view of Agrawal, et al. [1] and Information Builders Inc.

The Present Application:

For purposes of review, the present application relates to a method, an apparatus, and an article of manufacture providing for the creation and use of an index to access a subject multi-dimensional database. An embodiment of the invention receives an indication of a number of features of a subject multi-dimensional database to be identified, and then performs feature identification on the subject multi-dimensional database. An index for accessing the subject multi-dimensional database is created

9 of 18

Application No. 09/747,515
Amendment Dated June 1, 2004
Reply to Office Action of April 1, 2004

using the identified number of features. The index itself may be a multi-dimensional database.

As described in the specification on page 11, beginning at line 18:

In order to create the index, the Index System 124 passes the indexing parameters 128 to feature identification software 122 for use in finding "features" in the subject multi-dimensional database 136. Then, the Index System finds the "features" with the feature identification software 122. Next, the Index System builds the index 134 using the features as points for indexing into the subject multi-dimensional database 136. Then, the Index System 124 provides navigation capabilities for navigating the index 134 to access the subject multi-dimensional database 136. (emphasis added)

The invention provides a user interface to set up definitions for the subject multi-dimensional database to be mined, dimensions to be mined, measures to be mined, mining technique (i.e., feature identification) parameters, and number of results to be stored. The user interface directly drives a mining run. Additionally, the invention supports traversal of the multi-dimensional database, execution of the mining technique, and generation of result data. In one embodiment, the mining technique scans the subject multi-dimensional database and the result data is

Application No. 09/747,515
Amendment Dated June 1, 2004
Reply to Office Action of April 1, 2004

used to create an index. Operations for creating and deleting the index are provided as well as for outline definition, data population, cell note creation, and link partition definition. The invention provides capabilities for exploration and visualization of the result data against the subject multi-dimensional database.

Overall, the invention automatically builds the index, along with links to the subject multi-dimensional database. The invention also stores the index data in a spreadsheet data file, so that a spreadsheet user could view a list of deviations in one spreadsheet and link to cells in the subject multi-dimensional database using a linked partition mechanism.

There are many advantages to the invention including its straightforward implementation. Also, the invention does not require any additional functions or support from the developers of the subject multi-dimensional database, it does not modify the existing subject multi-dimensional database, and it does not store extra data in the subject multi-dimensional database. Extra explanations are selectively stored as cell notes on the index as desired, providing visualization and navigation of the multi-dimensional data. Furthermore, the invention is easily managed and can be applied with any data mining technique that can identify point of interest in a multi-dimensional database (i.e., a feature identification technique). The infrastructure of the invention supports plug-in techniques and can extend the solution beyond deviation detection.

Application No. 09/747,515
Amendment Dated June 1, 2004
Reply to Office Action of April 1, 2004

U.S. Patent No. 5,727,199 to Chen, et al.:

U.S. Patent No. 5,727,199 to Chen, et al. teaches a system, two-step method and program which, given a large training set of data tuples, in a first phase, called the feature identification phase, identifies features, which have good power in separating data tuples, based on a subset of the training set. In a second phase, called the feature combination phase, the identified features are evaluated in combination against the entire training set to determine final classification rules.

As described in column 1 of the Chen, et al. patent, various data mining capabilities have been explored in the literature and data mining is a broad field with many application-dependent problems requiring different mining techniques to solve. One application is "mining association rules" from a database of sales transactions. Another data mining application involves mining classification rules from large computer databases. As specified at line 45, the focus of the Chen, et al. patent is mining classification rules from large computer databases.

In column 2, it is described that a computer system is provided wherein operations of multi-feature extraction and development of classification rules from a large training database are performed more efficiently than previously known. Beginning at the bottom of that column, the method comprises first evaluating each feature in a subset of the training set as a function of its correlation with a group identifier, and identifying those features, evaluated as having an ID score exceeding a predetermined ID threshold, and, second, combining identified features

Application No. 09/747,515
Amendment Dated June 1, 2004
Reply to Office Action of April 1, 2004

into feature pairs and evaluating each feature pair in the training set as a function of its correlation with any one of the group identifiers, and identifying those feature pairs having a combination score which exceeds a predetermined feature combination threshold; and classifying the database as a function of the identified features and the identified feature pair.

Thus, it is respectfully submitted that the Chen, et al. patent does not teach, suggest, or fairly disclose using identified features in any way connected with creating an index for a subject multi-dimensional database. Rather, the identified features are used only to classify a database.

In addition to the above, it is respectfully submitted that the extraction and development of classification rules taught in the Chen, et al. patent, particularly in the feature identification phase, are not usable to create an index for a subject multi-dimensional database.

U.S. Patent No. 5,647,058 to Agrawal, et al.:

U.S. Patent No. 5,647,058 to Agrawal, et al. teaches a high dimensional indexing method which takes a set of objects that can be viewed as N-dimensional data vectors and builds an index which treats the objects like k-dimensional points. The method first defines and applies a set of feature extraction functions that admit some similarity measure for each of the stored objects in the database. A feature vector is then transformed in a manner such that the similarity measure is preserved and that the information of the feature vector is concentrated in only a

Application No. 09/747,515
Amendment Dated June 1, 2004
Reply to Office Action of April 1, 2004

few coefficients. The entries of the feature vectors are truncated such that the entries which contribute little on the average to the information of the transformed vectors are removed. An index based on the truncated feature vectors is subsequently built using a point access method (PAM).

It is respectfully submitted that the Agrawal, et al. [1] patent does not teach, suggest, or fairly disclose performing feature identification to identify an indicated number of features, and creating an index for a subject multi-dimensional database using the identified number of features.

U.S. Patent No. 6,094,651 to Agrawal, et al.:

U.S. Patent No. 6,094,651 to Agrawal, et al. teaches a method for locating data anomalies in a k-dimensional data cube that includes the steps of associating a surprise value with each cell of a data cube, and indicating a data anomaly when the surprise value associated with a cell exceeds a predetermined exception threshold.

All Pending Claims are Patentably Distinct and Unobvious Over the References of Record:

Referring once again to the Office Action in greater detail, claims 55-59, 61-65, and 67-71 were rejected as being unpatentable over U.S. Patent No. 5,727,199 to Chen, et al. (Chen) in view of U.S. Patent No. 5,647,058 to Agrawal et al. (Agrawal[1]). Claims 73-76, 78-81, and 83-86 were rejected as being unpatentable over U.S. Patent No. 5,727,199 to Chen, et al. (Chen) in view of

Application No. 09/747,515
Amendment Dated June 1, 2004
Reply to Office Action of April 1, 2004

U.S. Patent No. 5,647,058 to Agrawal et al. (Agrawal[1]) and further in view of Information Builders Inc. With regard to claims 55-59, 61-65, and 67-71, the Examiner took the position that Chen teaches a method, apparatus, and article of manufacture of accessing a subject multi-dimensional database stored on a data store connected to a computer, comprising:

- a) receiving an indication of a number of features of said subject multi-dimensional database to be identified (col. 3, lines 56-61 and col.4, lines 30-34);
- b) performing feature identification to identify the indicated number of features (col. 4, line 60 - col. 7, line 30)..."

It should be noted that Chen defines features as the n attributes of the database tuple $\{A_1, A_2, \dots, A_n\}$ (col. 3, lines 48-50). Features, as described in the present application, however, are more complex in nature. For example, an exemplary scenario is described starting on page 13, line 12 where the object is to obtain the three most prominent features. On page 14, lines 9-17, a table is provided illustrating the data returned by the feature identification technique of the particular embodiment being described. The result, as described in lines 1-4 on the same page, is an ordered list of multi-dimensional points, identifying regions of interest.

As can be seen from the above, the Chen patent does not teach, suggest, or fairly disclose receiving an indication of a number of features, and performing feature identification to identify the indicated number of features of the type described in the instant application.

Application No. 09/747,515
Amendment Dated June 1, 2004
Reply to Office Action of April 1, 2004

Applicants agree with the Examiner that "Chen, et al. does not clearly teach a step of creating an index for the subject multi-dimensional database using the identified number of features." As noted above, the results of the "feature identification phase" taught in the Chen, et al. patent is used (useful) for deriving mining classification rules from large computer databases but is not used (useful) for creating an index for a subject multi-dimensional database.

In addition to the above, the Agrawal, et al. patent [1] does not teach or suggest performing feature identification to identify an indicated number of features and creating an index for a subject multi-dimensional database using the identified number of features. Rather, as noted by the Examiner, the Agrawal, et al. patent [1] only teaches high dimensional indexing by taking N-dimensional data vectors and building an index of k-dimensional points.

Thus, the Chen, et al. and Agrawal, et al. [1] patents are not combinable in the manner suggested by the Examiner. In addition, even if those patents were somehow combined, they would not result in a method, apparatus, or article of manufacture capable of receiving an indication of a number of features to be identified, performing feature identification to identify the indicated number of features, and creating an index for the subject multi-dimensional database using the identified number of features.

For at least the above reasons, it is respectfully submitted that independent claims 55, 61, 67,

Application No. 09/747,515
Amendment Dated June 1, 2004
Reply to Office Action of April 1, 2004

73, 78, and 83 and dependent claims 56-59, 62-65, 68-71, 74-77, 79-82, and 84-86 depending respectively therefrom, are patentably distinct over the references of record.

Application No. 09/747,515
Amendment Dated June 1, 2004
Reply to Office Action of April 1, 2004

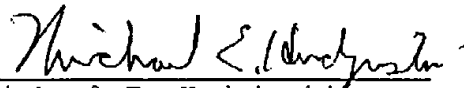
Conclusion

In view of the above amendments, comments, and arguments presented, applicants respectfully submit that all pending claims are patentably distinct and unobvious over the references of record.

Allowance of all claims and early notice to that effect is respectfully requested.

Respectfully submitted,

FAY, SHARPE, FAGAN,
MINNICH & MCKEE, LLP


Michael E. Hudzinski
Reg. No. 34,185
1100 Superior Avenue,
Seventh Floor
Cleveland, OH 44114-2518
(216) 861-5582 (tel)
(216) 241-1666 (fax)

N:\IBM\200022\GT\0000184V001.doc